Phylogenetics Analysis of Some Leguminosae Family Species Based On *RBCL* Sequence Data

Sagar S. Patel[*]**and Dipti B. Shah**

G. H. Patel Post Graduate Department of Computer Science and Technology,
Sardar Patel University, Vallabh Vidyanagar, Gujarat-388120, India.
[*]Corresponding author: Sagar Patel, Email: sgr308@gmail.com

ABSTRACT:

Phylogeny is the study of evolutionary relationships between species. Phylogeny is necessary to find out species' relationship with other species. In this paper, rbcL gene sequences data from NCBI database are considered for analysis. The sequence data of the rbcL gene are widely used in the reconstruction of phylogenies throughout the seed plants. Leguminosae family is one of the largest families that contain thousands of species of Plants, Herbs, Shrubs and Trees worldwide. This study shows that species of Leguminosae family which is further classified into Fabaceae (Papilionaceae), Mimosaceae and Caesalpiniaceae; based on morphological characters has different members and the based on the DNA and PROTEIN rbcL sequence data analysis, few species are not related with each other as per morphological classification. We conclude that as per botanical classification of species belonging to Leguminosae families are classified differently based on morphological character with compare to DNA and PROTEIN rbcL sequence data.

Keywords: Leguminosae family, Bioinformatics, NCBI, rbcL.

[I] INTRODUCTION

Leguminosae family contains species of Plants, Herbs, Shrubs, and Trees. Legumes are used as crops, forages and green manures; they also synthesize a wide range of natural products such as flavours, drugs, poisons and dyes. Legumes are able to convert atmospheric nitrogen into nitrogenous compounds useful to plants. This is achieved by the presence of root nodules containing bacteria of the genus Rhizobium. These bacteria have a symbiotic relationship with Legumes, fixing free nitrogen for the plants; in return legumes supply the bacteria with a source of fixed carbon produced by photosynthesis. This enables many legumes to survive and compete effectively in nitrogen poor conditions. Leguminosae family is further classified into three subfamilies; Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaceae.

1.1. RbcL gene

The sequence data of the rbcL gene are widely used in the reconstruction of phylogenies throughout the seed plants. Comprehensive phylogenetic analyses of Leguminosae began with the plastid gene rbcL (Doyle, 1995; Ka¨ss and Wink, 1995, 1996; Doyle et al., 1997) following the early, widespread use of this gene for phylogenetic studies of land plant relationships (e.g., Chase et al., 1993).

1.2. NCBI (The National Center for Biotechnology Information)

The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI houses a series of databases relevant to biotechnology and biomedicine. Major databases include GenBank for DNA sequences, Protein, Genome, EST etc. All these databases are available online through the Entrez search engine.

1.3. DNA (Deoxyribonucleic acid) / Nucleotide

The Deoxyribonucleic acid (DNA) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. Along with RNA and proteins, DNA is one of the three major macromolecules essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information.

1.4. Protein

Proteins are large biological molecules consisting of one or more chains of amino acids. Proteins perform a vast array of functions within living organisms, including catalyzing metabolic reactions, replicating DNA, responding to stimuli, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in folding of the protein into a specific three-dimensional structure that determines its activity.

[II] MATERIALS AND METHODS

In this paper we have considered around 266 species which are found in Gujarat state of India. Further we searched each species in NCBI database and finally found around 148 species' information like DNA, Protein and other useful information of leguminosae family. Further we have only considered rbcL DNA and Protein sequences. Analysis done in MEGA software and analysis was done with Maximum Likelihood Method with Bootstrap method (figure 1).
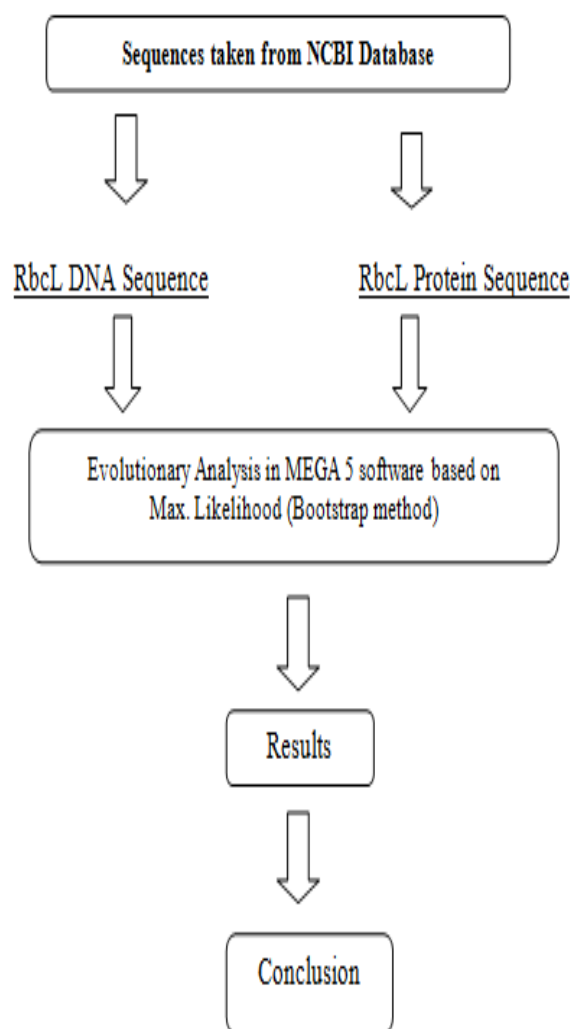

Figure 1. Flow chart of method.

[III] RESULTS
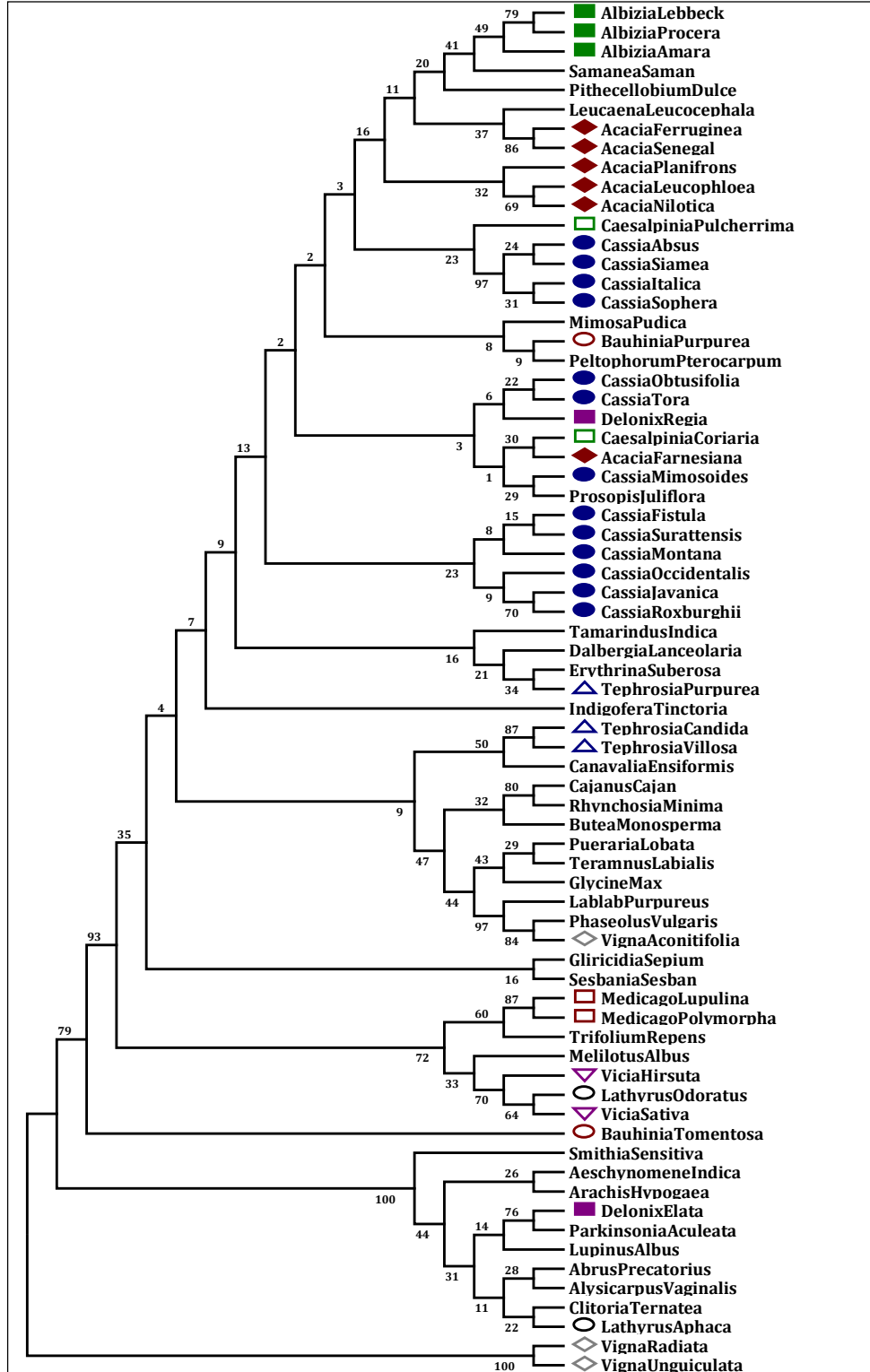## A. RESULTS OF *DNA RBCL* GENE SEQUENCES

Figure 2. Result of Max.Likelihood (Bootstrap Method) of rbcL DNA sequences of Leguminosae Family

As shown in above figure 2 which is result of DNA rbcL Gene Sequences by Max.Likelihood method (bootstrap method), starting from top we observed that species are placed in subfamily wise; like first Mimosaceae, Caesalpiniaceae followed by Fabaceae (Papilionaceae). Species of genus *Albizia* and *Acacia* genus are related as per morphological characters or botanical classifications except *Mimosa pudica, Acacia farnesiana* and *Prosopis juliflora* are found in Caesalpiniaceae subfamily as these species might be homologues with nearby species of Caesalpiniaceae subfamily.

Then after from *Caesalpinia pulcherrima* to *Tamarindus indica* these species are belongs to Caesalpiniaceae subfamily; in that few species of *Cassia* genus are related as per morphological characters or botanical classifications and species of *Caesalpinia, Delonix* and *Bauhinia*

genus are distantly related with each other. Further, species like *Delonix elata, Bauhinia tomentosa* and *Parkinsonia aculeata* are found in Fabaceae subfamily.

Species from *Dalbergia lanceolaria* to *Vigna unguiculata* belongs to Fabaceae subfamily. In this subfamily species of *Tephrosia, Medicago, Vicia and Vigna* are related as per morphological characters or botanical classifications. However species of *Lathyrus* genus are distantly related with each other. None of single species of this subfamily has it's species in other species. This shows that this subfamily's species are closely related with each other and as per the botanical or morphological classification this subfamily is more accurate than other subfamily of Leguminosae Family.

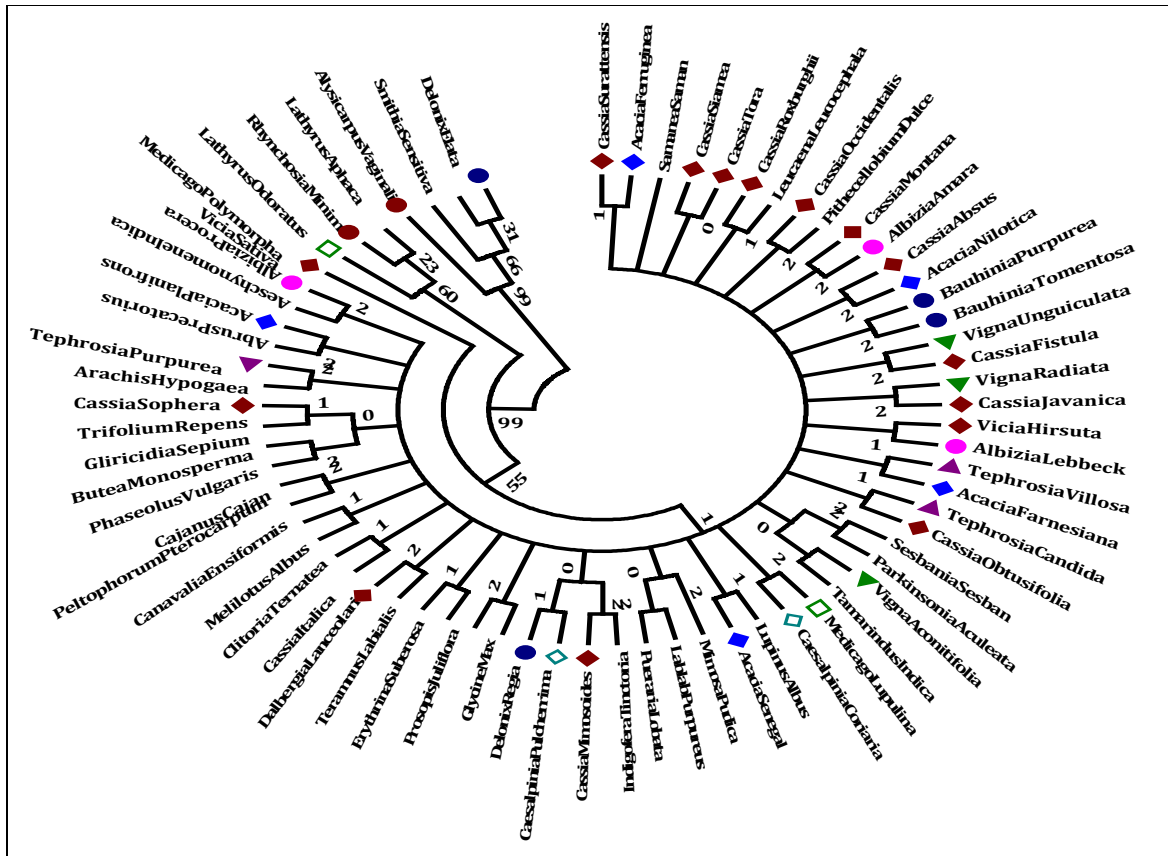## B. RESULTS OF *PROTEIN RBCL* GENE SEQUENCES

Figure 3. Result of Max.Likelihood (Bootstrap Method) of rbcL Protein sequences of Leguminosae Family

As shown in above figure 3 which is result of Protein rbcL Gene Sequences by Max.Likelihood method (bootstrap method), starting from top we observed that species are placed in subfamily wise; like first, Caesalpiniaceae, Mimosaceae followed by Fabaceae (Papilionaceae). This figure is in circular form in which shows that species from each subfamily is related with each other in very complex manner. In this result what we found that species of *Cassia* and *Acacia* genus are distributed in other each subfamily and species of *Bauhinia* and few species of *Cassia* genus are which indicates that as per the related as per morphological characters or botanical classifications but species of *Albizia, Vigna, Vicia, Tephrosia, Medicago, Caesalpinia, Lathyrus and Delonix* are distantly related with each other which is not true in respect of botanical or morphological classification.

[IV] CONCLUSION

In this study we observed that species belongs to Leguminosae Family; which is further classified into Fabaceae (Papilionaceae), Mimosaceae, Caesalpiniaceae are as per the botanical classification classified differently based on their morphological features like species' flower color, size and shape, types and arrangements of Stipules, size of plant etc. But this study focus on evolutionary relationship of Leguminosae Family species based on DNA & Protein sequences of rbcL gene with Multiple sequence alignment by Maximum likelihood where we observed that in rbcL DNA sequences result; some species belonging to same genus are fall very nearly as per botanical classification which is correct as per both botanical and evolutionary relationship but we observed in rbcL PROTEIN sequence result that as compare with rbcL DNA sequence it really differs and it is not related with morphologically or botanical classification and further we observed that few species are distantly related even if they are from same genus. So, we

conclude that rbcL DNA sequence analysis is strongly related with species' botanical or morphological classification but result of rbcL PROTEIN sequence is not related with species botanical or morphological classification. We suggest considering rbcL DNA sequences are better for evolution study for small group of species.

ACKNOWLEDGEMENT

REFERENCES

[1] G. L. Shah (1978): Flora of Gujarat State. Publ. by Sardar Patel University, Vallabh Vidyanagar, Anand, India.

[2] Kalpesh Anjaria (2002) Ph. D. Thesis: Floristic studies of Anand District. Submitted to Sardar Patel University, Vallabh Vidyanagar, Anand, India.

[3] Sagar Patel (2011) Some Leguminous trees in Anand District (M.Sc., Project work) Sardar Patel University, Vallabh Vidyanagar, Gujarat, India.

[4] Jean-Mchel Claverie and Cedric Notredame (2003) Bioinformatics – A Beginner's Guide. Publ. by Wiley Publishing, Inc. USA.

[5] G. M. Oza; Kishore S. Rajput (2006) Biodiversity of Gujarat Forest Trees.Publ. By INSONA, Vadodara, India.

[6] Patel, Anjaria, Panchal (2012) Leguminous Trees In Anand District: Collection and Analysis With Bioinformatics Applications. LAP LAMBERT Academic Publishing, Germany.

[7] Sagar Patel, Hetalkumar Panchal. Bioinformatics information of Leguminosae family in Gujarat state in International Journal of Agriculture Environment & Biotechnology. Vol 7, Issue1, March 2014, Pages: 11-15. ISSN 0974-1712. Online ISSN 2230-732X.

[8] Sagar Patel, Panchal H., 2013. Leguminobase: A Tool To Get Information Of Some Leguminosae

Family Members From Ncbi Database in Journal of Advanced Bioinformatics Applications and Research: Vol 4, Issue3, 2013, Pages. 54-59. ISSN 0976-2604. Online ISSN 2278 – 6007.

[9] Sagar Patel, Panchal H., Smart J., Anjaria K., 2013. Species Information Retrieval Tool: A Bioinformatics tool for Leguminosae family in International Journal of Bioinformatics and Biological Science: Vol.1 n.2 Pages.187-194 June, 2013 Print ISSN 2319-5169

[10] Sagar Patel, Panchal H., Smart J., Anjaria K., 2013. Distribution of Leguminosae family members in Gujarat State of India: Bioinformatics Approach in International Journal of Computer Science and Management Research, Pages- 2184-2189 Vol 2 Issue 4 April 2013, ISSN 2278-733X

[11] Sagar Patel, Panchal H., Anjaria K., Phylogenetic analysis of some leguminous trees using CLUSTALW2 Bioinformatics Tool. Publication Year: 2012, Page(s): 917- 921. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, USA, Oct 4-7, 2012; E-ISBN: 978-1-4673-2744-2; Print ISBN: 978-1-4673-2746-6.

[12] Sagar Patel, Panchal H., Anjaria K., DNA Sequence analysis by ORF FINDER & GENOMATIX Tool: Bioinformatics Analysis of some tree species of Leguminosae Family. Publication Year: 2012, Page(s): 922- 926. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, USA, Oct 4-7, 2012; E-ISBN: 978-1-4673-2744-2; Print ISBN: 978-1-4673-2746-6.

[13] Harborne, J.B. 1994. Phytochemistry of the Leguminosae. In Phytochemical Dictionary of the Leguminosae, eds Bisby,F.A. et al. London: Chapman & Hall

[14] Martin F. Wojciechowski,matt Lavin,michael J. Sanderson. A Phylogeny Of Legumes (Leguminosae) Based On Analysis Of The Plastid Matk Gene Resolves Many Well-supported Subclades Within The Family.

[15] Polhill, R.M. & Raven, P.H. (eds) 1981. Advances in Legume Systematics. Royal Botanic Gardens, Kew

[16] Smartt, J. & Simmonds, N.W. (eds) 1995. Evolution of Crop Plants. Harlow: Longman Scientific & Technical

[17] Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 (Tamura, Peterson, Stecher, Nei, and Kumar 2011).

[18] Hayashi K. Kawano S. 2000. Molecular systematics of Lilium and allied genera (Liliaceae): phylogenetic relationships among Lilium and related genera based on the rbcL and matK gene sequence data. Plant Species Biology 15: 73-93.

[19] Hilu K. W. Borsch T. Müller K. Soltis D. E. Soltis P. S. Savolainen V. Chase M. W. Powell M. P. Alice L. A. Evans R. Sauquet H. Neinhuis C. Slotta T. A. B. Jens G. R. Campbell C. S. Chatrou L. W. 2003. Angiosperm phylogeny based on matK sequence information. American Journal of Botany 90: 1758-1776.

[20] Hilu KW, Liang H: The matK gene: sequence variation and application in plant systematics. American Journal of Botany 1997, 84:830-839.

[21] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution (submitted).

[22] http://www.en.wikipedia.org

[23] http://www.theplantlist.org/browse/A/Leguminosae

[24] http://www.ildis.org/

[25] http://www.ncbi.nlm.nih.gov/

[26] http://www.kew.org/

[27] http://www.missouribotanicalgarden.org/

[28] http://plantnet.rbgsyd.nsw.gov.au/iopi/iopihome.htm

[29] http://www.ncbi.nlm.nih.gov/nuccore/

[30] http://www.ncbi.nlm.nih.gov/protein/